

ABSTRACT OF THE DISCLOSURE

A method of distributing workload in a workflow management system includes the step of calculating a load index for each engine of the workflow management system. The load index is calculated as an average activity execution delay attributable to the associated engine. The load indices indicate average execution latency between the start of consecutive activity nodes of a process due to engine loading exclusive of resource execution time. Workload is distributed across the plurality of engines in a load sensitive mode. In one embodiment, distribution switches from a load insensitive mode to a load sensitive mode for distributing processes when a maximum differential load index exceeds a first pre-determined threshold. Distribution switches from the load sensitive mode back to the load insensitive workload distribution mode for distributing processes when the maximum differential load index is less than a second pre-determined threshold.